A Reservoir Model for Intra-Sentential Code Switching Comprehension in French and English

Pauline Detraz
Inria, Bordeaux, France

Xavier Hinaut
Inria, Bordeaux, France

Abstract

Some people can mix two languages within the same sentence: this is known as intra-sentential code-switching. The majority of computational models on language comprehension are dedicated to one language. Some bilingual models have also been developed, but very few have explored the code-switching case. We collected data from human subjects that were required to mix pairs of given sentences in French and English. Truly bilingual subjects produced more switches within the same sentence. The corpus obtained have some very complex mixed sentences: there can be until eleven language switches within the same sentence. Then, we trained ResPars, a Reservoir-based sentence Parsing model, with the collected corpus. This Recurrent Neural Network model processes sentences incrementally, word by word, and outputs the sentence meaning (i.e. thematic roles). Surprisingly the model is able to learn and generalize on the mixed corpus with performances nearly as good as the unmixed French-English corpus.